



Performance Prediction: Where the Rubber Meets the Road

Adolfy Hoisie

**Parallel Architectures and Performance Team
Modeling, Algorithms and Informatics CCS-3**

**with: Eitan Frachtenberg, Darren Kerbyson,
Mike Lang, Scott Pakin, Fabrizio Petrini, Harvey
Wasserman, and others**

Work funded by ASCI/Institutes

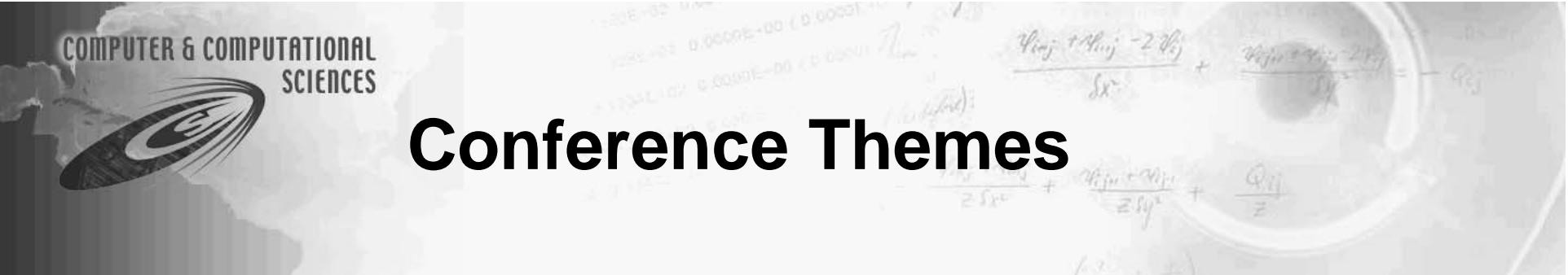
Computer and Computational Sciences Division
hoisie@lanl.gov
www.c3.lanl.gov





Outline

- Overview of Performance Modeling
- Applications of Performance Models



Conference Themes

- What is the best balance of HW components to solve an app in the fastest wall-clock time?
- How much effort should we expend tuning a code for improved performance and how much we expect to gain?



Performance

- “DOE is peak performance oriented”
- “Big iron vs. clusters”: isn’t big iron a glorified cluster nowadays?
- Metrics: Linpack vs. other (“superior”) single number performance
- Resource requirements for ASCI codes: e.g. bandwidth



Performance Analysis

- Measuring performance is of limited use:
 - ◆ current implementation of code
 - ◆ currently available architectures
 - ◆ impossible to distinguish between real performance and machine idiosyncrasies
- Design space is Multidimensional
 - ◆ runtime = f (microprocessor performance, memory hierarchy, network characteristics, compiler/language etc.)
- Performance Characterization
 - ◆ typically done based on cross-sections of the design space

Performance is a Multidimensional Space

- problem size: surface-to-volume
- # of processors: scalability
- architectural design (size, topology, etc)
- communication parameters
- computation parameters
- optimal (problem) blocking sizes
- target optimization (e.g., runtime, problem size)



Performance Analysis Methods

- Analytical (Los Alamos, ORNL, IBM)
- Statistical (Wisconsin, UIUC, NERSC)
 - ◆ System workload performance
 - ◆ Mean Value Analysis
- Queuing theory
- Experimental
 - ◆ Simulation (UCLA, Dartmouth, Los Alamos)
 - ◆ Benchmarking (UT, Los Alamos, ORNL, Purdue)
 - ◆ Trace-driven experiments (UIUC, Barcelona)



Selection of Performance Analysis Method...

“ More than any other time in history, mankind faces a cross-roads. One path leads to despair and utter hopelessness. The other, to total extinction. Let us pray we have the wisdom to choose correctly ”

- *Woody Allen*



Performance Engineering

- Performance-engineered system: The components (application and system) are **parameterized** and **modeled**, and **constitutive model is proposed** and **validated**.
- **Predictions** are made based on the model. The model is meant to be **updated**, **refined**, and **further validated** as new factors come into play.



Analytical Performance Modeling

- An “analytical” representation of the “fundamental equation of modeling”

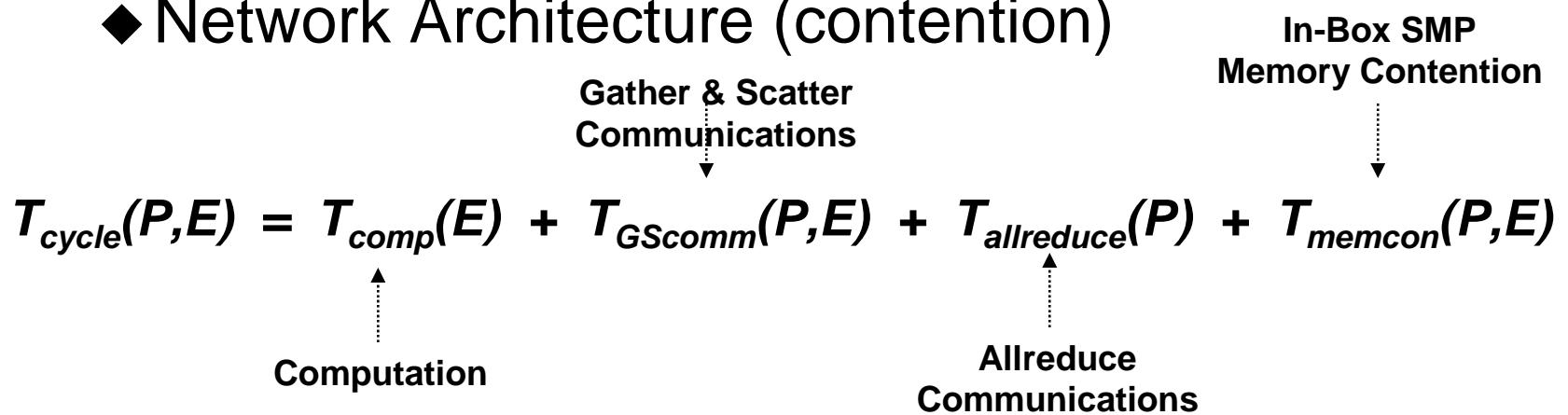
$$T_{\text{runtime}} = T_{\text{comm}} + T_{\text{comp}} - T_{\text{overlap}}$$

for given application

- A model captures the algorithm-architecture mapping
- A model is parametric
- A model is seldom analytically “clean”

Analytical Performance Modeling

- Encapsulates code characteristics
- Parameterized in terms of:
 - ◆ Code parameters (number of cells)
 - ◆ System parameters (CPU speed, communication latency & bandwidth, memory contention)
 - ◆ Network Architecture (contention)





Analytical Performance Modeling

- A tradeoff between model complexity and accuracy applies for a significant part of the application spectrum (mesh refinement, non-deterministic, etc)
- A methodology for analytical modeling of entire applications developed at Los Alamos



State-of-the-art/affairs

at Los Alamos:

- Sn transport on structured grids

Adolfy Hoisie, Olaf Lubeck, Harvey Wasserman, Fabrizio Petrini, Hank Alme, A General Predictive Performance Model for Wavefront Algorithms on Clusters of SMPs, LAUR-00308, In the proceeding of ICPP 2000, August 20-25, 2000. Toronto, Canada.

- Hydrodynamics (CSD): Sage

Darren J. Kerbyson, Hank J. Alme, Adolfy Hoisie, Fabrizio Petrini, Harvey J. Wasserman, Michael Gittings, "Predictive Performance and Scalability Modeling of a Large-Scale Application," in proceedings of SC2001.

- Sn transport on unstructured grids
- Non-deterministic transport (Monte-Carlo)

elsewhere:

- A model of a climate code: "Performance Modeling for SPMD Message-Passing Programs", Concurrency: Practice and Experience, April 1998, Brehm, Worley and Madhukar.
- A model of an MD code: "Demonstrating the Scalability of a MD Application on a Petaflop Computer", Almasi et al.

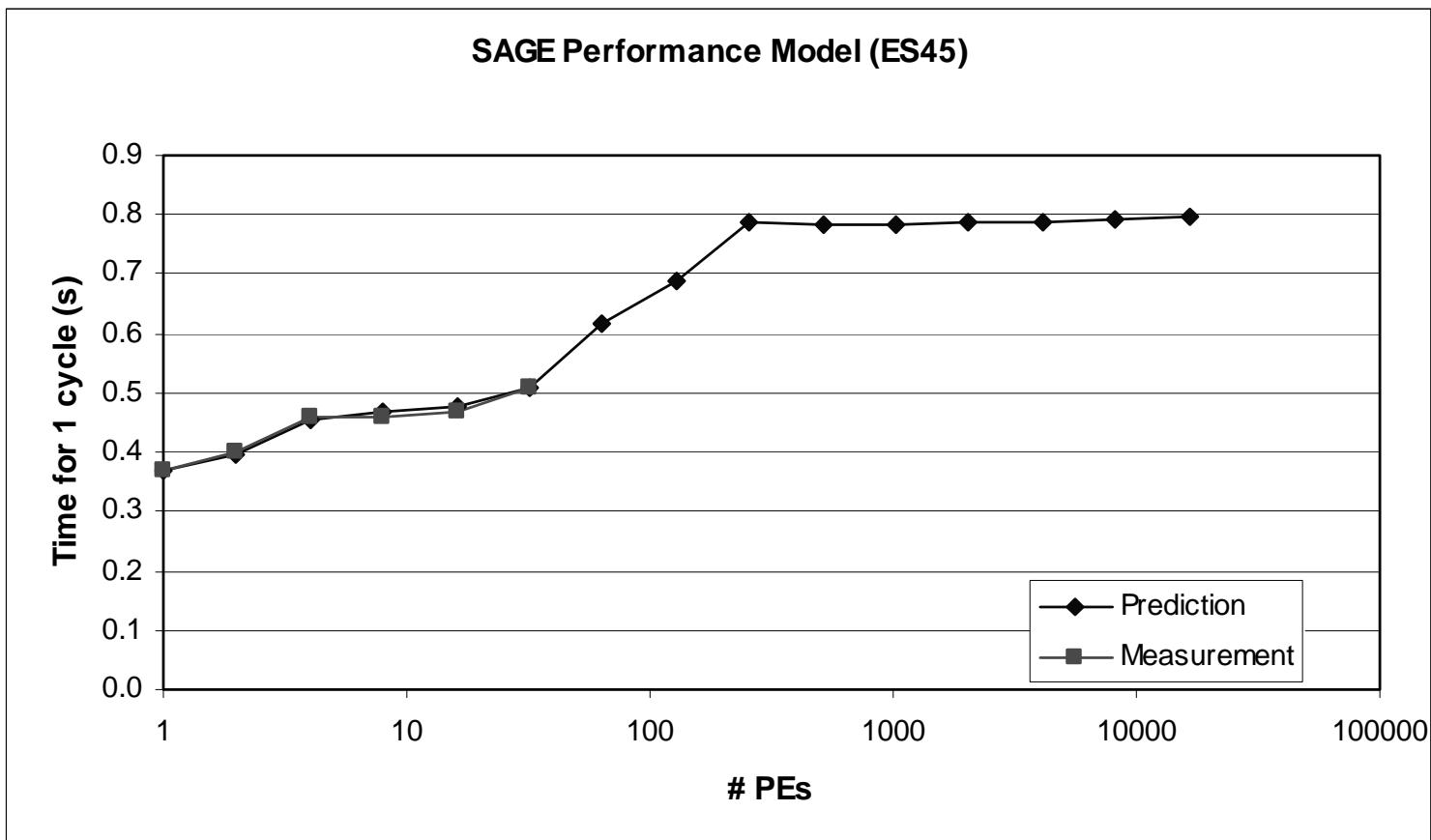


Predictive Value of Performance Models

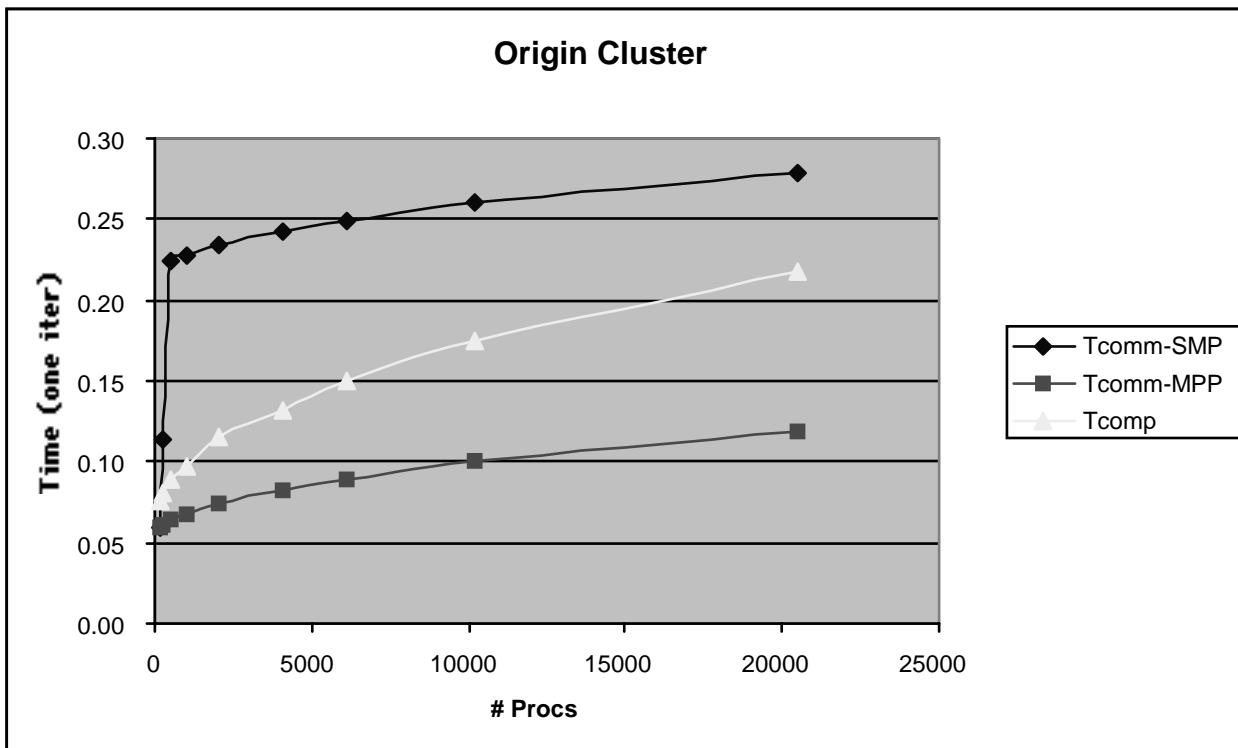
- No need for implementation
- Fast exploration of the design space
- Use model to explore:
 - ◆ Change of existing Architectures
 - ☞ e.g. change in sub-system performance (increased communication bandwidth, upgrade of CPU etc)
 - ◆ Future Architectures (non-existing)
 - ☞ Compare alternatives
- Will describe specific/actual uses for (I) architecture and application design exploration, and (II) System Diagnostics

Architecture exploration (I)

- Predictions for the 30T

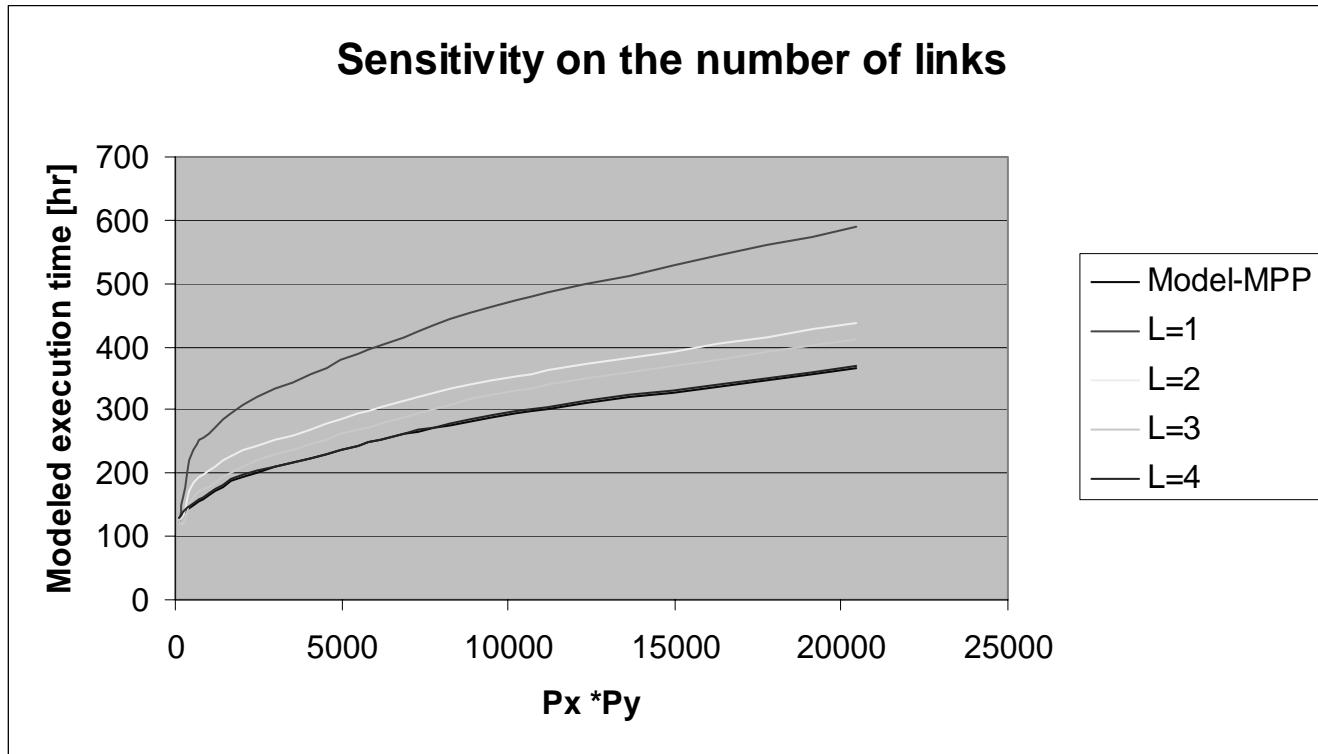


Architecture exploration (II)



50 MFLOPS/cpu, L=100 us, Bw=100MB/s,
4 x 4 x 100 subgrid, optimal blocking, 10e7 cells total,
1 Link ea. Dir. Between Hosts.

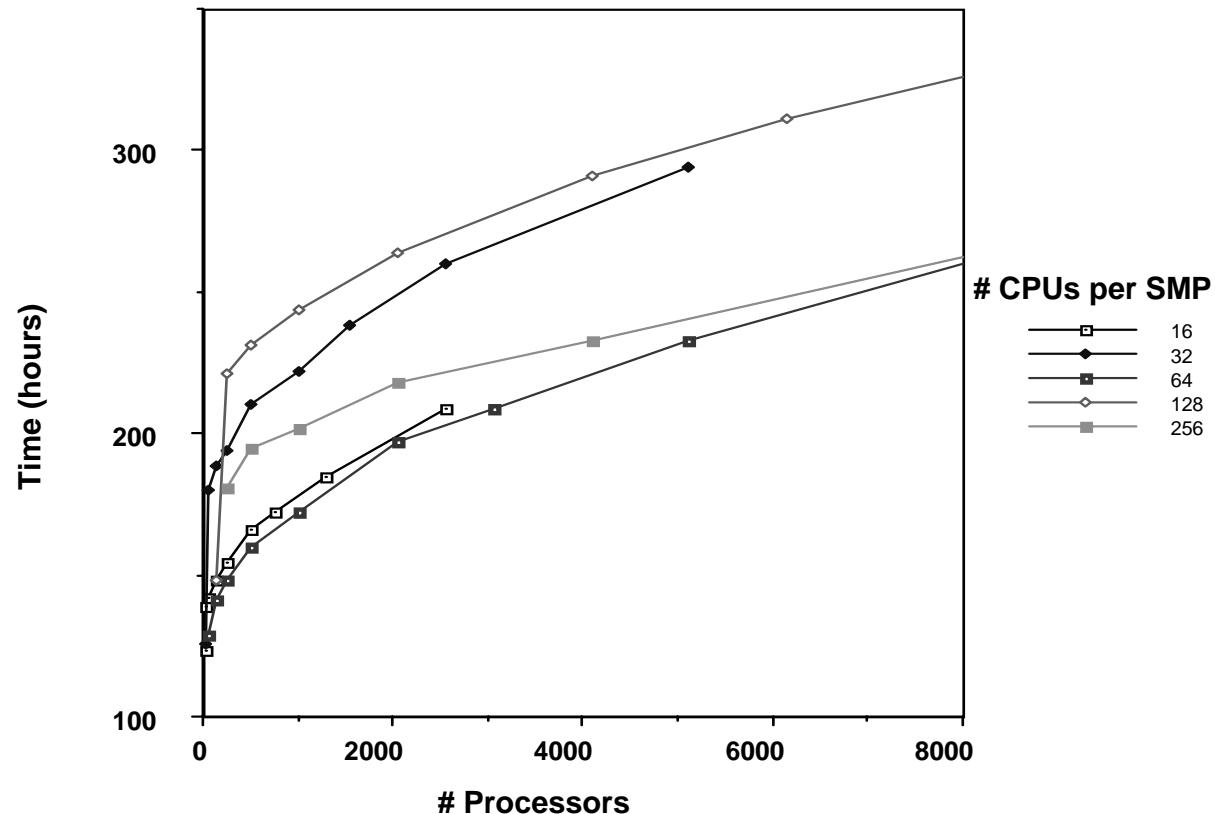
Architecture (II)



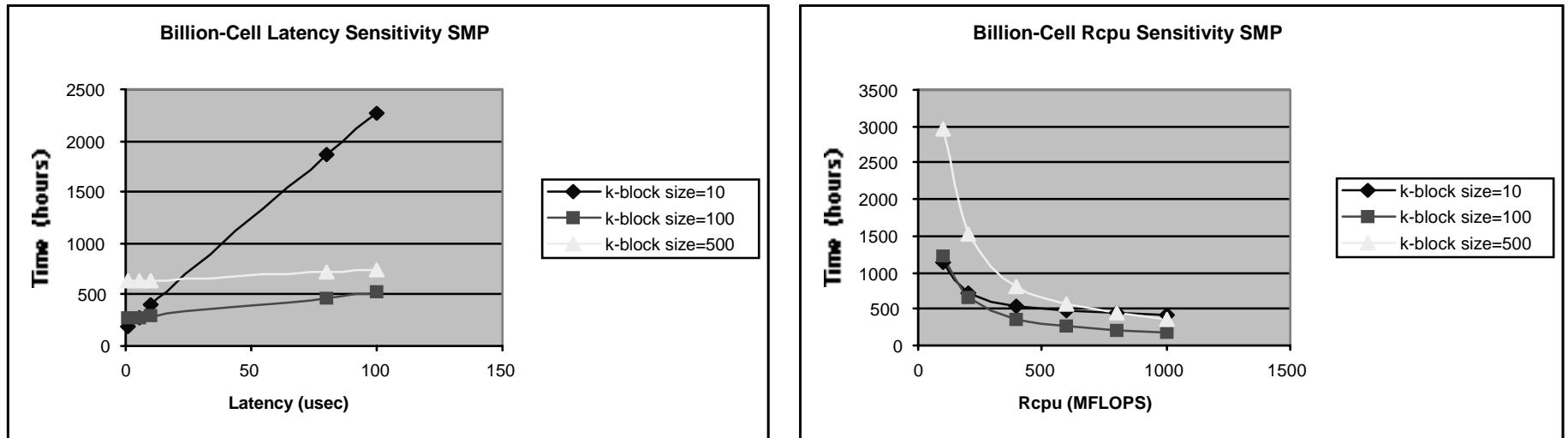
50 MFLOPS/cpu, L=100 us, BW=100MB/s,
4 x 4 x 100 subgrid, optimal blocking, 10e7 cells total,
NG=30, 12 iters, 10e4 timesteps

Architecture exploration (III): Sensitivity Analysis on SMP Size

$L = \min(s_x, s_y)/4$



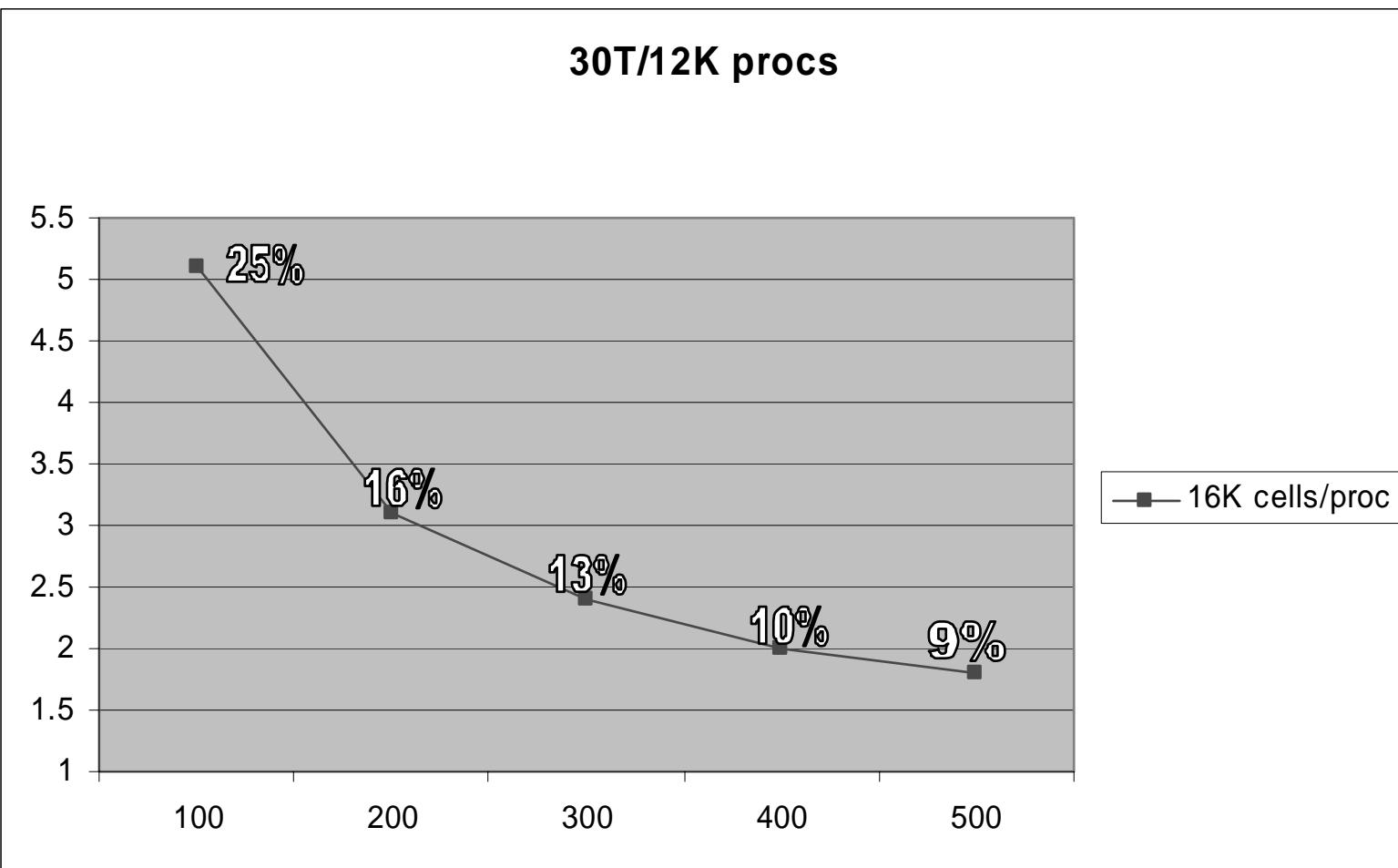
Architecture exploration (IV) Particle Transport Scalability Results: 1 Billion Cells on an SMP cluster



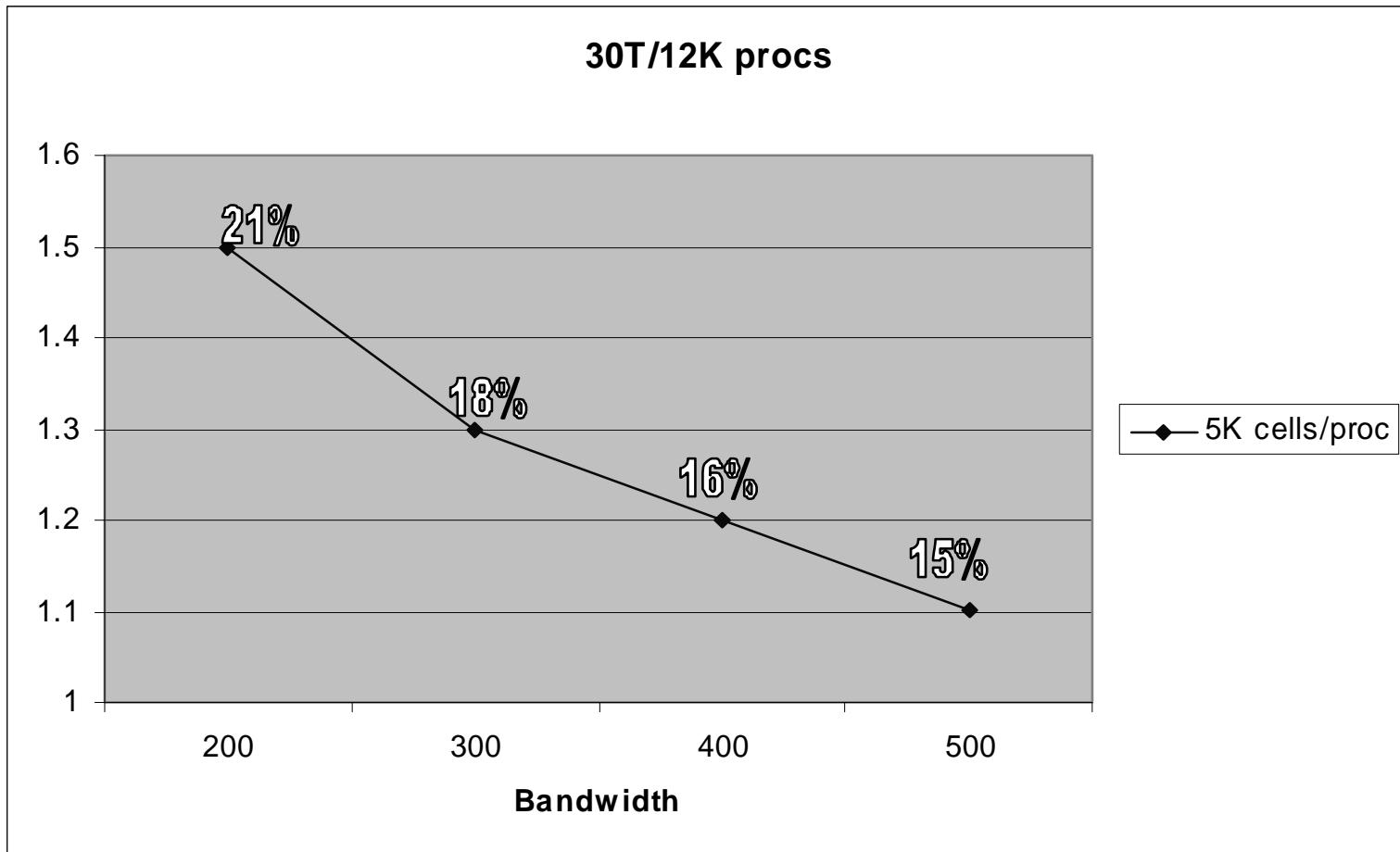
Estimates of SWEEP3D Performance on a Hypothetical Future-Generation (100-TFLOPS) System as a Function of MPI Latency and Sustained Per-Processor Computing Rate.

	Sustained Computing Rate	
	10% of Peak	50% of Peak
MPI Latency	Runtime (hours)	Runtime (hours)
0.1 μ s	178	58
1.0 μ s	205	68
10 μ s	264	104

Architecture exploration (V)

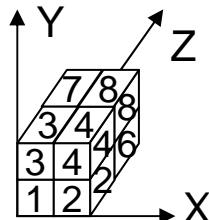


Architecture exploration (VI)

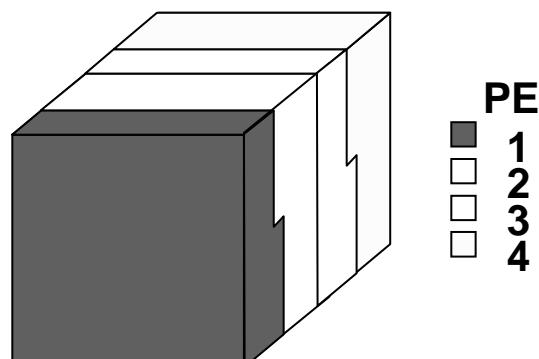


Application Optimization: Sage

- Assignment to processors done in blocks (2x2x2):



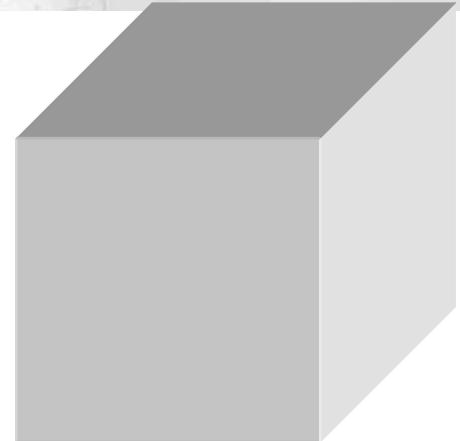
- “Distorted” slab decomposition





Application optimization

- 3-D cell Grid
- Partition in 3 dimensions,
 - ◆ each PE can have:
 i cells in X, j cells in Y, k cells in Z



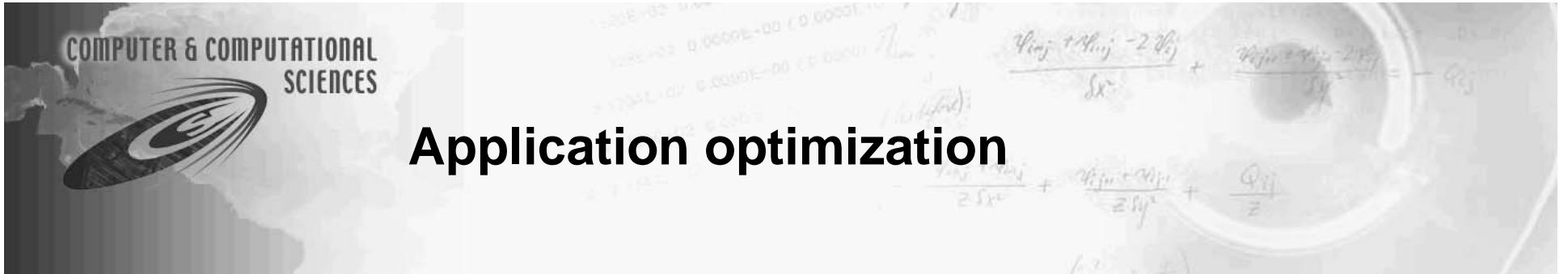
Volume = $i \cdot j \cdot k$ (computation)

Surface = $i \cdot j + j \cdot k + k \cdot i$ (communication - gather/scatter)

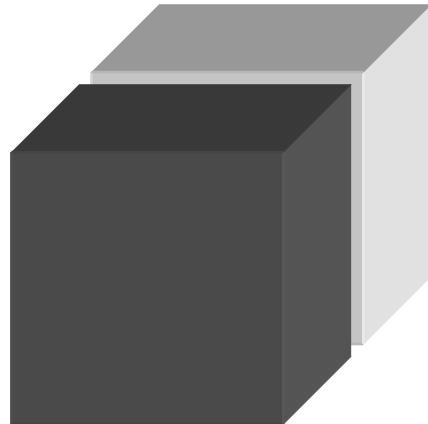
$$\frac{\text{Surface}}{\text{Volume}} = \frac{1}{i} + \frac{1}{j} + \frac{1}{k}$$

(min when $i=j=k$)

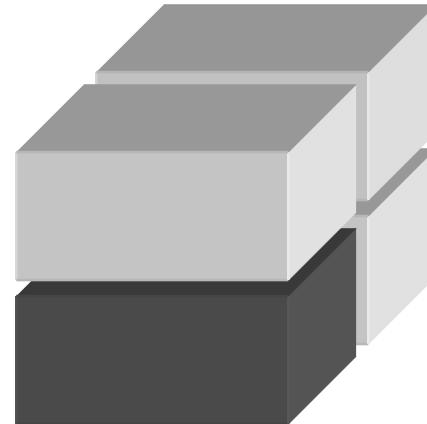
Adolfy Hoisie
Salishan, April, 2002



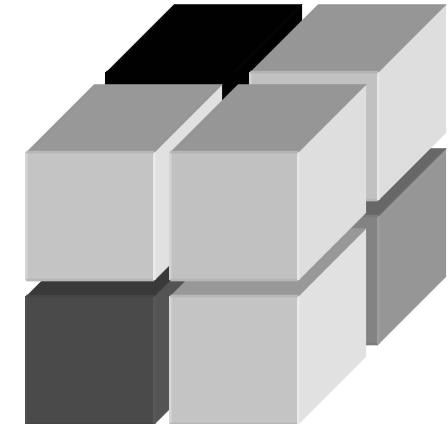
Application optimization



Case 1: 2x2x1



Case 2: 2x1x1

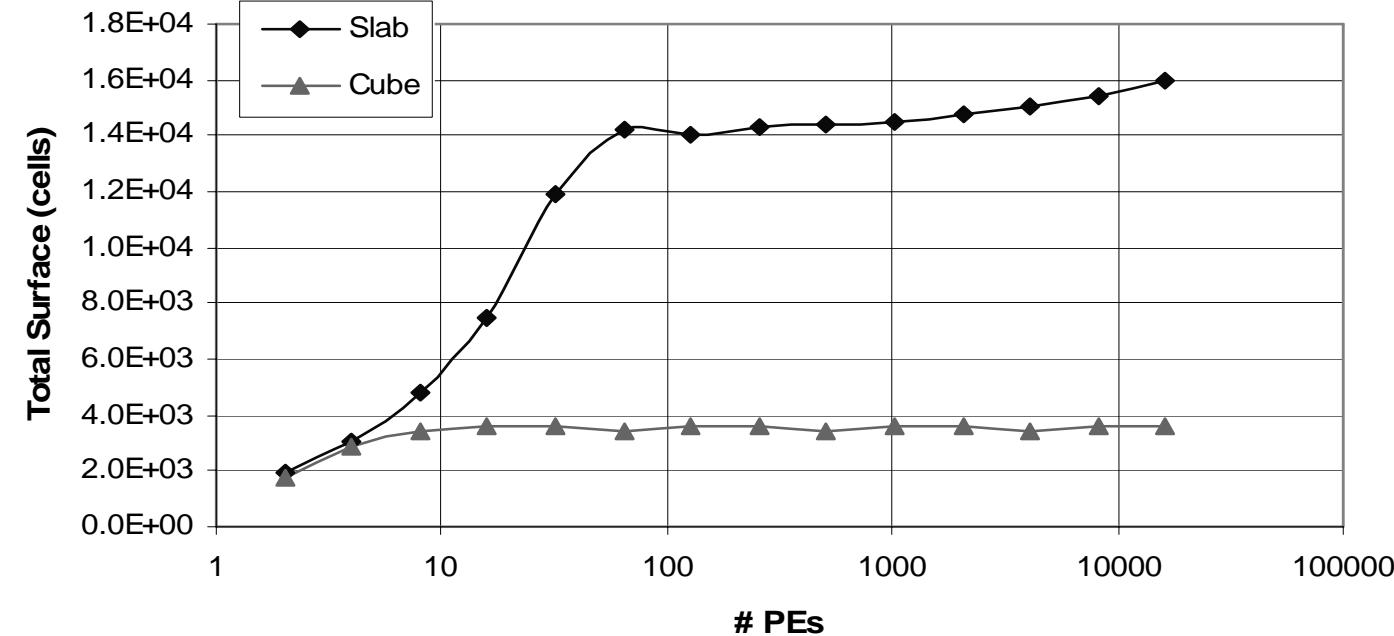


Case 3: 1x1x1

- Minimum Surface-to-Volume ratio
 - ◆ minimizes communication time (Gather & Scatter)

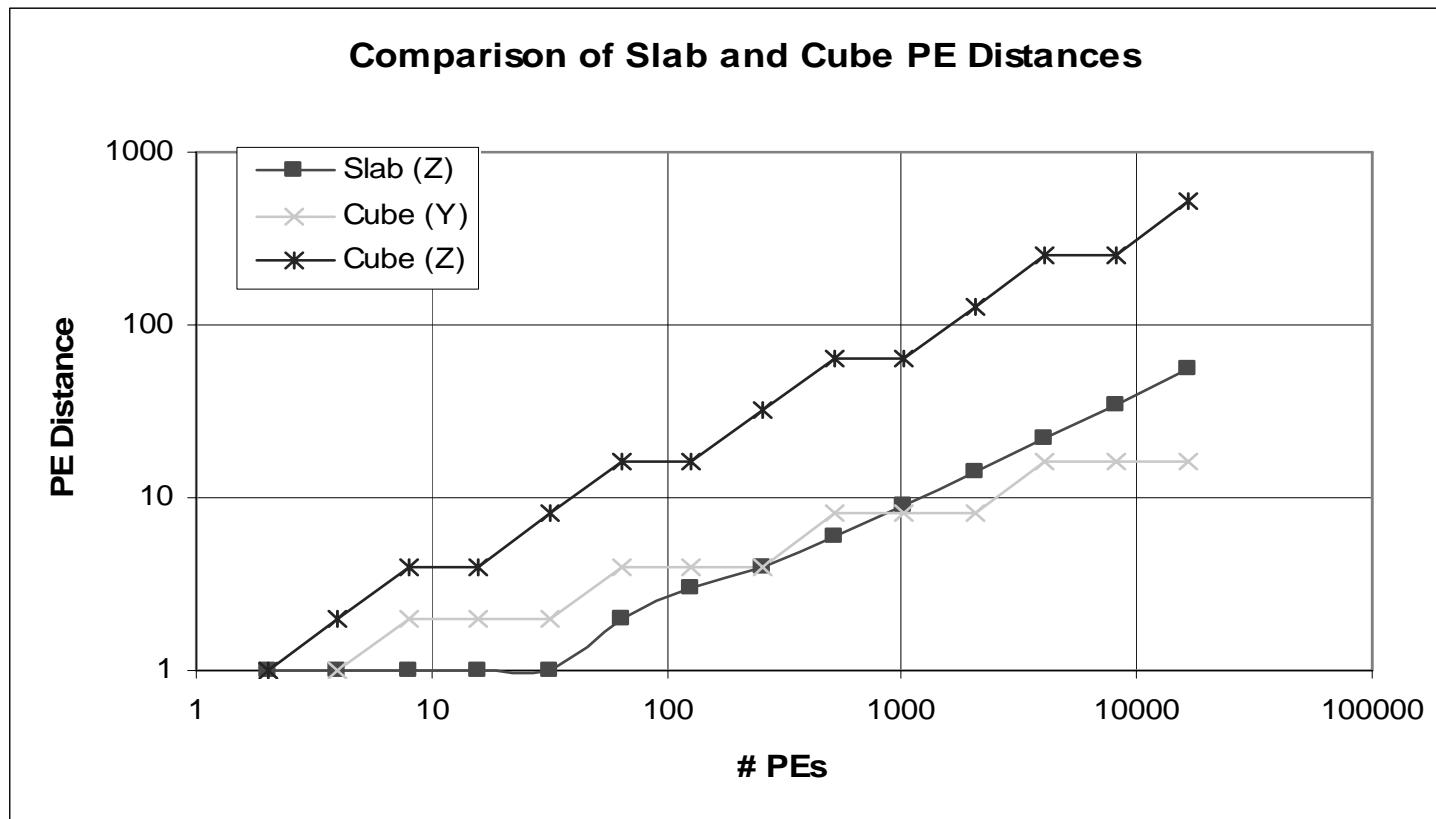
Application optimization Cube vs Slab (Compaq ES45)

Comparison of Slab and Cube PE Surface Sizes



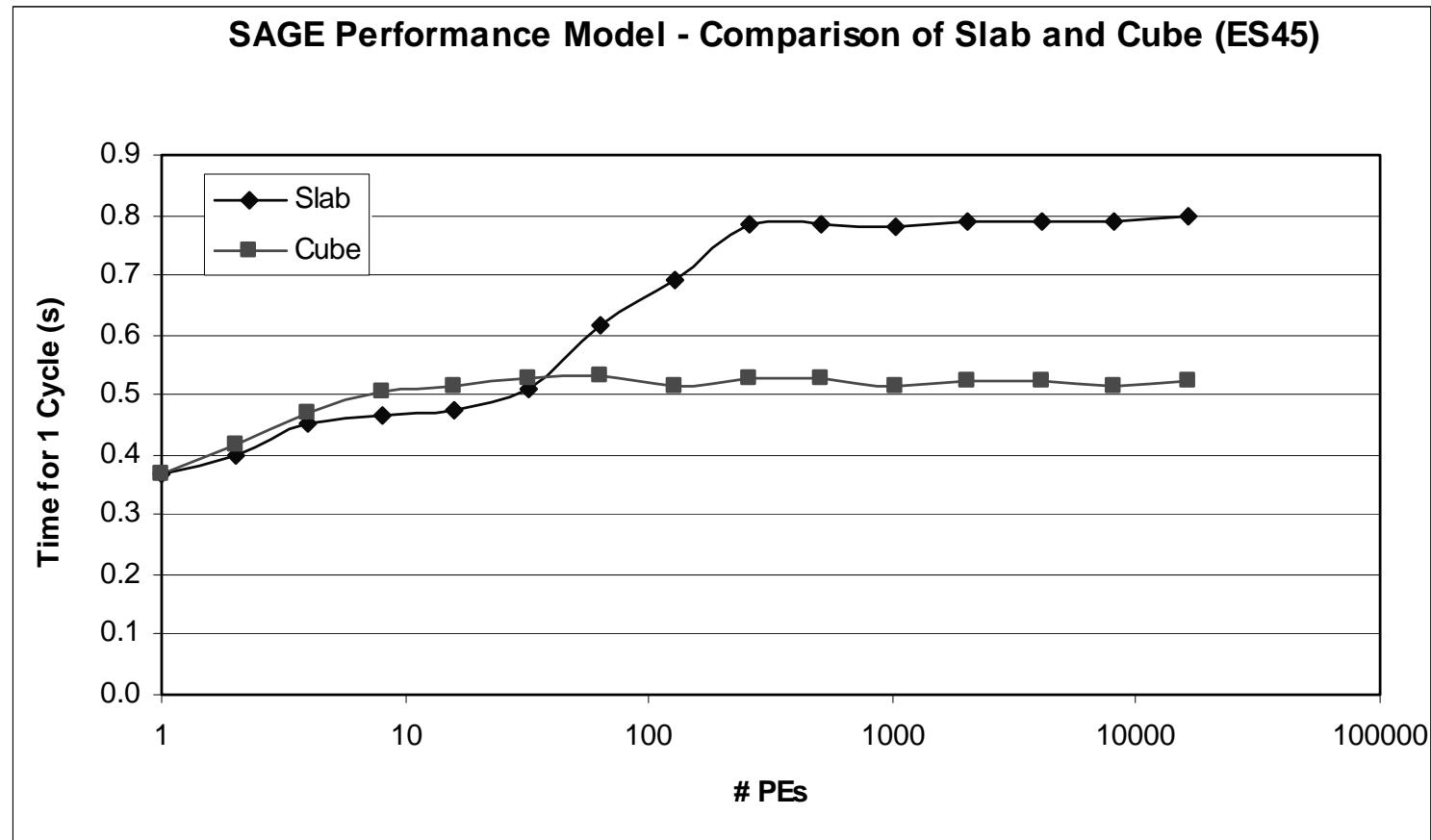
- Cube Surface: 4 times smaller than Slab

Application optimization Cube vs Slab (Compaq ES45)



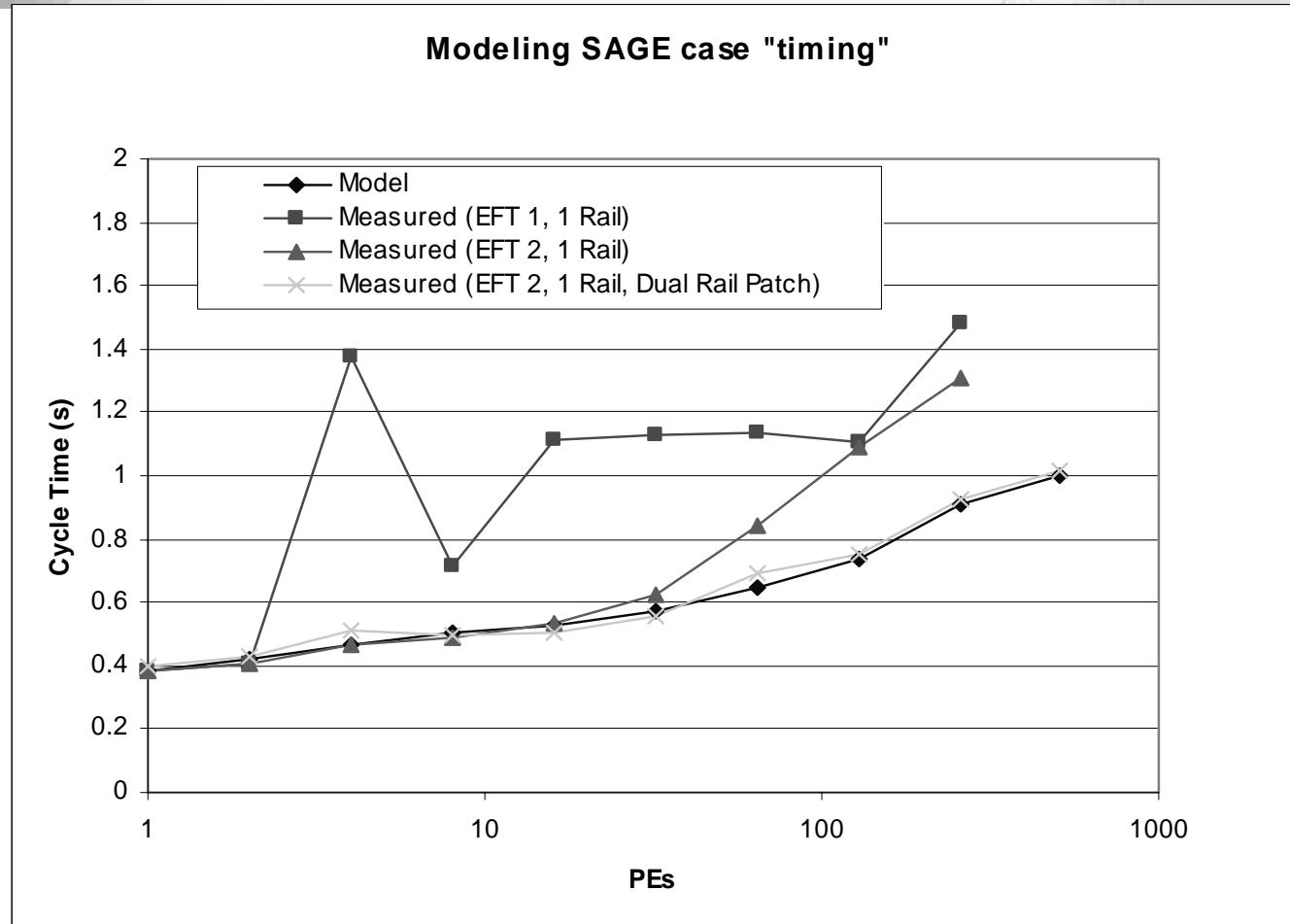
- Cube PE distance > Slab PE distance

Application optimization Cube vs Slab (Compaq ES45)



- Expect performance improvement using cube

Rational system integration





Future Research Directions

- Complete the workload characterization of the Los Alamos ASCI application workload, through modeling
- Expand the modeling effort to include non-ASCI apps
- Transform the static models into dynamic performance tools
- Application steering
- Performance Specification Languages (PSL) for describing performance models.



Conclusions

- Application / architecture mapping is the key - not lists of raw basic machine characteristics.
- Point design studies need to address a specific workload.
- Performance and scalability modeling is an effective “tool” for workload characterization, system design, application optimization, and algorithm-architecture mapping.
- Back-of-the-envelope performance predictions are risky (outright wrong ?), given the complexity of analysis in a multidimensional performance space.
- Applications and systems at this scale need to be performance-engineered -- modeling is the means to analysis.

Resources

Parallel Architectures and Performance Team
from http://www.c3.lanl.gov/par_arch

- **PAPERS:**

- **Fabrizio Petrini, Wu-chun Feng, Adolfy Hoisie, Salvador Coll, and Eitan Frachtenberg.** The Quadrics Network (QsNet): High-Performance Clustering Technology. In *IEEE Micro*, 22(1):46-57, January-February 2002
- **Darren J. Kerbyson, Hank J. Alme, Adolfy Hoisie, Fabrizio Petrini, Harvey J. Wasserman, Michael Gittings,** "Predictive Performance and Scalability Modeling of a Large-Scale Application," in proceedings of SC2001.
- **Fabrizio Petrini, Salvador Coll, Eitan Frachtenberg and Adolfy Hoisie.** Hardware- and Software-Based Collective Communication on the Quadrics Network. In *IEEE International Symposium on Network Computing and Applications 2001 (NCA 2001)*, Boston, MA, October 2001.
- **Salvador Coll, Eitan Frachtenberg, Fabrizio Petrini, Adolfy Hoisie, and Leonid Gurvits.** Using Multirail Networks in High-Performance Clusters. In *IEEE Cluster 2001*, Newport Beach, CA, October 2001.
- **Eitan Frachtenberg, Fabrizio Petrini, Salvador Coll and Wu-chun Feng.** Gang Scheduling with Lightweight User-Level Communication . In *2001 International Conference on Parallel Processing (ICPP2001), Workshop on Scheduling and Resource Management for Cluster Computing*, Valencia Spain, September 2001
- **Adolfy Hoisie, Olaf Lubeck, Harvey Wasserman, Fabrizio Petrini, Hank Alme,** A General Predictive Performance Model for Wavefront Algorithms on Clusters of SMPs, LAUR-00308, In the proceeding of ICPP 2000, August 20-25, 2000. Toronto, Canada

- **BOOK:**

- **Stefan Goedecker and Adolfy Hoisie,** "Performance Optimization of Numerically Intensive Codes (Software, Environments, Tools), Paperback - 173 pages (March 19, 2001) Society for Industrial & Applied Mathematics; ISBN: 0898714842.

The End

